






Article

Rockfall Source Identification Using a Hybrid Gaussian Mixture-Ensemble Machine Learning Model and LiDAR Data

Ali Mutar Fanos¹⁾ · Biswajeet Pradhan ^{2)†} · Shattri Mansor ³⁾ ·
Zainuddin Md Yusoff ⁴⁾ · Ahmad Fikri bin Abdullah ⁵⁾ · Hyung-Sup Jung ⁶⁾

Abstract: The availability of high-resolution laser scanning data and advanced machine learning algorithms has enabled an accurate potential rockfall source identification. However, the presence of other mass movements, such as landslides within the same region of interest, poses additional challenges to this task. Thus, this research presents a method based on an integration of Gaussian mixture model (GMM) and ensemble artificial neural network (bagging ANN [BANN]) for automatic detection of potential rockfall sources at Kinta Valley area, Malaysia. The GMM was utilised to determine slope angle thresholds of various geomorphological units. Different algorithms (ANN, support vector machine [SVM] and k nearest neighbour [k NN]) were individually tested with various ensemble models (bagging, voting and boosting). Grid search method was adopted to optimise the hyperparameters of the investigated base models. The proposed model achieves excellent results with success and prediction accuracies at 95% and 94%, respectively. In addition, this technique has achieved excellent accuracies (ROC = 95%) over other methods used. Moreover, the proposed model has achieved the optimal prediction accuracies (92%) on the basis of testing data, thereby indicating that the model can be generalised and replicated in different regions, and the proposed method can be applied to various landslide studies.

Key Words: Rockfall source identification, Gaussian mixture model, ensemble machine learning, GIS, LiDAR

1. Introduction

Landslides are categorized into different types based on the moving materials and the motion mode (Varnes,

1978). Consequently, rockfall refers to the moving materials are rocks and the motion mode is falling while landslide refers to the moving material is soil and the motion mode is sliding. In other words, rockfall

Received February 4, 2019; Accepted February 13, 2019; Published online February 19, 2019

¹⁾ PhD Scholar, Department of Civil Engineering, Universiti Putra Malaysia

²⁾ Distinguished Professor, Faculty of Engineering and Information Technology, University of Technology Sydney

³⁾ Professor, Department of Civil Engineering, Universiti Putra Malaysia

⁴⁾ Senior Lecturer, Department of Civil Engineering, Universiti Putra Malaysia

⁵⁾ Senior Lecturer, Department of Biological and Agricultural Engineering, Universiti Putra Malaysia

⁶⁾ Professor, Department of Geoinformatics, University of Seoul

† Corresponding Author: Biswajeet Pradhan (Biswajeet.Pradhan@uts.edu.au)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

is a subtype of landslide. Rockfall is a common natural hazard in many places worldwide, including Malaysia; this phenomenon affects transportation ways, infrastructure, urban areas and other socioeconomic activities located near steep rock slopes (Pradhan and Fanos, 2017). In particular, rockfall risks are increasing in mountainous regions given economic activities and population growth (Budetta, 2004). Rockfall is defined as a single block detached from a slope by falling, bouncing, rolling or/and sliding moving downslope (Varnes, 1978). These events can cause severe casualties because rockfall is difficult to be predicted and can move with a high velocity depending on the geometric and geomorphologic characteristics of the moving mass (Corona *et al.*, 2013).

Numerous studies, including identification of potential source areas (Yang *et al.*, 2017; Muzzillo *et al.*, 2018; Mote *et al.*, 2019), susceptibility mapping (Gigli *et al.*, 2014), trajectory analysis (Fanos *et al.*, 2016; Pellicani *et al.*, 2016; Budetta *et al.*, 2016; Pradhan and Fanos, 2017), risk assessment (Corona *et al.*, 2017; Mineo *et al.*, 2018; Moos *et al.*, 2018) and modelling of rock bounce heights and velocity (Fanos *et al.*, 2018) have been conducted on rockfall.

In particular, rockfall source identification is fundamental because it governs the rockfall run-out. This phenomenon is an important element in assessing rockfall probability and hazard. The identification of rockfall source areas can be performed through field investigation or inventory dataset of rockfall incidents. Nevertheless, such techniques are time-consuming and costly (Malamud *et al.*, 2004). In addition, inventory dataset is typically incomplete or lacking in time and space (Loye *et al.*, 2009). Many techniques have been developed recently for rockfall source identification considering accurate 3D terrain models and GIS dataset (Jaboyedoff and Labiouse, 2003; Gigli *et al.*, 2014). The general concept of the existing techniques is based on identification slope angle thresholds which are considered unsteady. For example, Jaboyedoff and

Labiouse (2003) and Guzzetti *et al.* (2003) used thresholds $>45^\circ$ and $>60^\circ$, correspondingly. Furthermore, Acosta *et al.* (2007) proposed an advanced method on the basis of the geometry of a slope derived from LiDAR dataset and other conditioning factors that utilise data mining, statistical and probabilistic methods. Agliardi *et al.* (2016) aimed to identify unstable rocks using a terrestrial photogrammetric technique in composite-structure regions. Messenzehl *et al.* (2017) evaluated various conditioning factors that control rockfall at a regional scale. Their result showed that rockfall is controlled by various conditioning factors with different relative importance.

Although the aforementioned studies have exerted notable efforts to propose techniques for identifying potential rockfall sources using photogrammetric or LiDAR data, one major issue remains unsolved. That is, the area of interest which contains other types of landslides that have nearly comparable conditioning factors, such as shallow landslide and rockfall. Consequently, it is difficult to differentiate rockfall from other landslide types based on just the probability map. Therefore, additional factor that can contribute to this problem has to be used and assessed. However, Fanos *et al.* (2018) used an individual machine learning algorithm to differentiate various landslide types. Nevertheless, in their study, the hyperparameters of the used algorithm which highly affect the realistic of the obtained results were not optimized. In addition, they employed a limited conditioning factors without testing the multicollinearity among these factors and optimization. Moreover, they applied GMM based on the inventory dataset not on the geomorphological units of the slope. Consequently, the current research proposes a hybrid model for identifying potential rockfall sources by utilising airborne laser scanning dataset and other conditioning factors considering the gaps mentioned above. The proposed hybrid model combines two main approaches (i.e. Gaussian mixture model [GMM] and bagging artificial neural networks

[BANN]). The proposed model is implemented and evaluated on three datasets in Ipoh, where several types of landslides have occurred. The key motivation of this research is to use the generated maps in order to avoid more urbanization in hazardous areas and have a sustainable environment. In addition, the produced information can decrease the requirement to perform in-situ investigation and the identified source areas can be used to carry-out further assessment of rockfall hazard and risk.

2. Characteristics of the Study Area

Ipoh is selected as the study area because it is situated within the Perak state in Peninsular Malaysia (Fig. 1).

The city is surrounded by Keledang from the west, Tambun from the east, Chemor from the north and Kampung Kepayang from the south. Ipoh is approximately 220 km north of Kuala Lumpur (capital of Malaysia). Geographically, the study area is situated between the northeast ($101^{\circ}8'30''$, $4^{\circ}39'00''$) and the southwest ($101^{\circ}3'30''$, $4^{\circ}31'30''$) corners. The major land-use features include urban, tin-mining and non-operational areas, oil palm plantation forest, shrubs, peat swamp forest and grassland.

Ipoh experiences tropical climate with temperature ranging from 25°C to 35°C throughout the year with comparatively rising humidity (approximately 82.3%) (Meteorological Service Department of Malaysia). Ipoh encounters intensive rainfall, excluding the dry season (May, June and July). In addition, the city

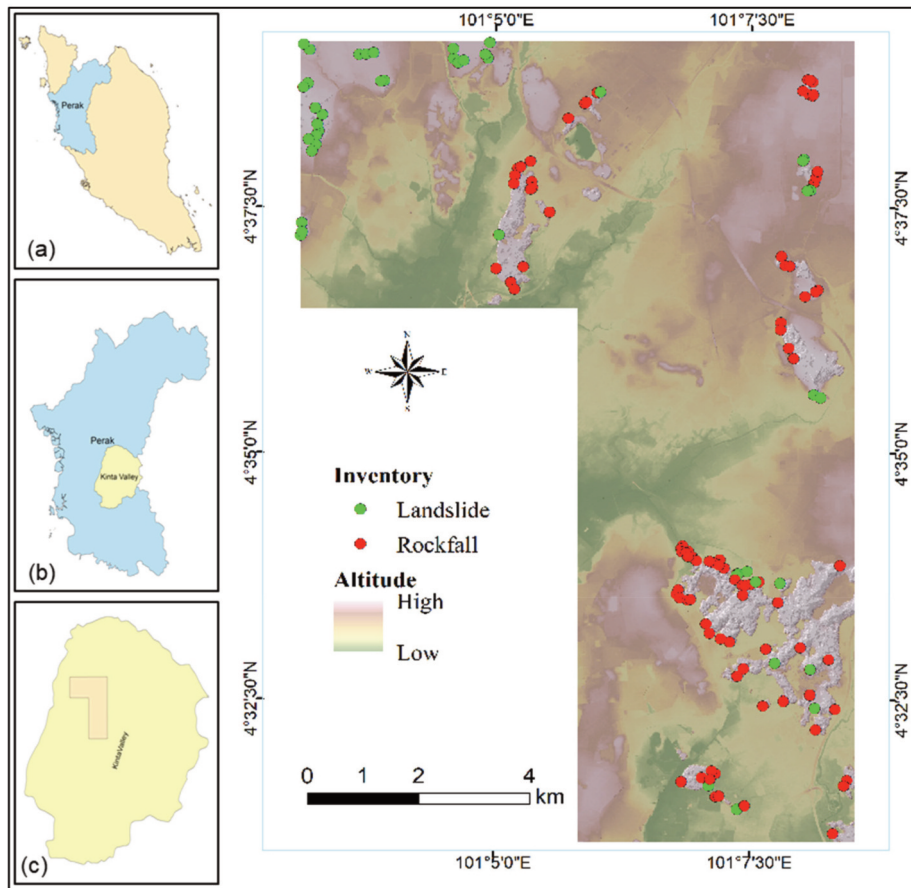


Fig. 1. Study area (a) Malaysia, (b) Perak State, (c) Kinta Valley, and (d) Ipoh.

receives an average rainfall of 319 mm annually (Malaysia Meteorological Service Department).

Geologically, the study area consists of diverse lithology with a wide presence of igneous rocks. Such features typically exist in regions with high altitude on the east and west sides of Ipoh. In addition, sedimentary (limestone) and metamorphic (marble) rocks are profusely present in the study area. Notably, Ipoh is situated within Sunda Shield plate. According to a continental plate study, this plate separates and moves away annually by 10 mm towards the east of the Eurasian plate (Pradhan *et al.*, 2014).

3. Materials and Methods

This section presents an overview of the used datasets and the proposed integrated model for identifying potential rockfall sources using the LiDAR dataset. The proposed model is based on the GMM and ensemble ANN method called BANN. The details of this model and its implementation and validation methods are presented in the following subsections. The GMM was implemented using MATLAB (2017), whilst the ensemble models were implemented through Python.

1) Datasets

The main dataset used in this research is LiDAR data and landslide inventory. This section describes these datasets and the derived conditioning factors.

(1) Description of LiDAR Data

The laser scanning dataset was obtained using an airborne LiDAR sensor specified with a frequency rate of 25,000 Hz and flight height of 1,500 m in 2016. Consequently, high-density point clouds were collected with approximately 10 pts/m². A point cloud of approximately 800 million of data samples were obtained. The acquired raw data were subjected to

pre-processing to eliminate noises and outliers. In addition, a filtering process was conducted through GIS environment to separate non-ground points from ground points. An interpolation method (Inverse Distance Weighted) was employed to produce the DTM based on ground points. Thus, accurate DTMs were generated and utilised to derive the landslide conditioning factors.

(2) Landslide and Rockfall Inventory Map

Landslide inventory data are a key element in landslide probability modelling for two purposes (i.e. to train and then validate the model). Various sources were used to prepare the landslide inventory dataset that involves historical records (Department of Mineral and Geoscience Malaysia), field measurements and remote sensing. High-resolution SPOT fused images and aerial photo (0.1 m) were utilised for the visual inspection of landslides within the focus region. However, several landslide events can occur within regions that are invisible in the satellite image or beneath vegetation. Therefore, such events were collected through historical records and field measurements (for old and new events, respectively). Multiple in situ campaigns were performed using a precise global navigation satellite system (GNSS). The surveys included the whole area of interest especially the areas that previously reported in the literature on the recorded reports. This to map the unrecorded or new incidents. In addition, a verification process was carried-out to verify the recorded incidents. Consequently, the locations of fresh landslide scars were specified and mapped through a field survey. A total of 147 samples with their related properties were prepared for landslide assessment (Fig. 1). These landslides are shallow landslide and rockfall. The statistical analysis of the inventory dataset shows that the shallow landslides were occurred within slope angles range from 19° to 53°, while rockfalls were occurred within slope angles range from 51° to 79°. Regarding the lithology, almost all the landslide

incidents were located within the limestone area. The prepared inventory dataset was split into two subsets (i.e. training [70%] and testing [30%] of the data samples) to ensure that the datasets encompass all the landslide types by stratified sampling (Hong *et al.*, 2016). Both training and testing data samples were selected randomly and each subset contains landslide and non-landslide samples. The training data were used for training the models, and the remaining data were utilised for model optimisation (10%) and accuracy testing (20%).

(3) Conditioning Factors

The raw LiDAR dataset contains up-ground and ground points, and therefore, the up-ground features must be removed using a filtering algorithm to produce an accurate DTM that represents the bare earth surface. On this basis, multiscale curvature (MCC) algorithm was used in this research for LiDAR data filtering (Evans and Hudak, 2007). Consequently, the DTM was generated on the basis of the remaining points using inverse weighted distance (IDW) interpolation method. The data statistics revealed a vertical accuracy of 0.15 m (root mean square error) and a horizontal accuracy of 0.3 m.

Rockfalls are controlled by various conditioning factors, and each factor has different relative importance (Pourghasemi *et al.*, 2018). Therefore, this research utilises several conditioning factors for identifying the rockfall sources in the presence of shallow landslide in Ipoh. These components encompassed morphological, vegetation, anthropogenic, lithological and hydrological factors (Fig. 2). These factors are commonly listed in literature as they affect the strength of terrain, the possibility of triggering by climate, earthquake, and earth movements, and the potential erosion, flow direction and length, and deposition. Multicollinearity among these factors was assessed to remove the insignificant factors that can adversely affect the performance of the proposed model by

increasing the complexity and variation in the conditioning factors. Therefore, the coefficient of determination (R^2) was determined and then the variance inflation factor (VIF) was calculated for each factor. Factors with VIF of more than 4 were considered insignificant and may adversely influence the modelling thus such factors should be removed.

The morphological factors (i.e. slope, altitude, aspect and curvature) were extracted from the generated DTM and GIS spatial analysis tools. These factors were produced as continuous raster files. Slope is the elevation change rate in the direction of the steepest descent. This factor is one of the main factors that control landslides. The altitude factor is usually controlled by many geomorphological and geological processes. For instance, landslide often occurs at moderate elevation because slope tends to be covered by a thin colluvium layer, that is prone to landslide. Moreover, aspect is the direction of slope from the north in a clockwise direction and it ranges from 0° to 360° . The curvature influences the convergence and divergence of flow along a surface, the acceleration and deceleration of downslope flows and, thus, affects erosion and deposition. This factor was calculated using the second derivative of the DTM. This factor. Moreover, flow length and distances to stream, road and lineaments were included. The flow length affects the rock falling process that influences the runout distance and energy loss. The intermittent flow regime of gullies and hydrological network encompasses saturation and erosive processes, thereby increasing pore water pressure and resulting in a landslide in regions close to drainage channels. Vegetation elimination, extensive excavation and creation of road networks are common processes in slopes. Moreover, lineaments are regarded as the main prompting factor for landslides. Buffers (closeness) to these features increase the landslide probability.

The anthropogenic factors included land use. The land use map was produced by classifying SPOT 5

satellite images and high-resolution aerial photo (0.1 m). In addition, in situ investigation was performed for verifying the land use map. The lithology of Ipoh is

mainly marble/limestone in addition to sandstone and granite.

Furthermore, four hydrological factors included

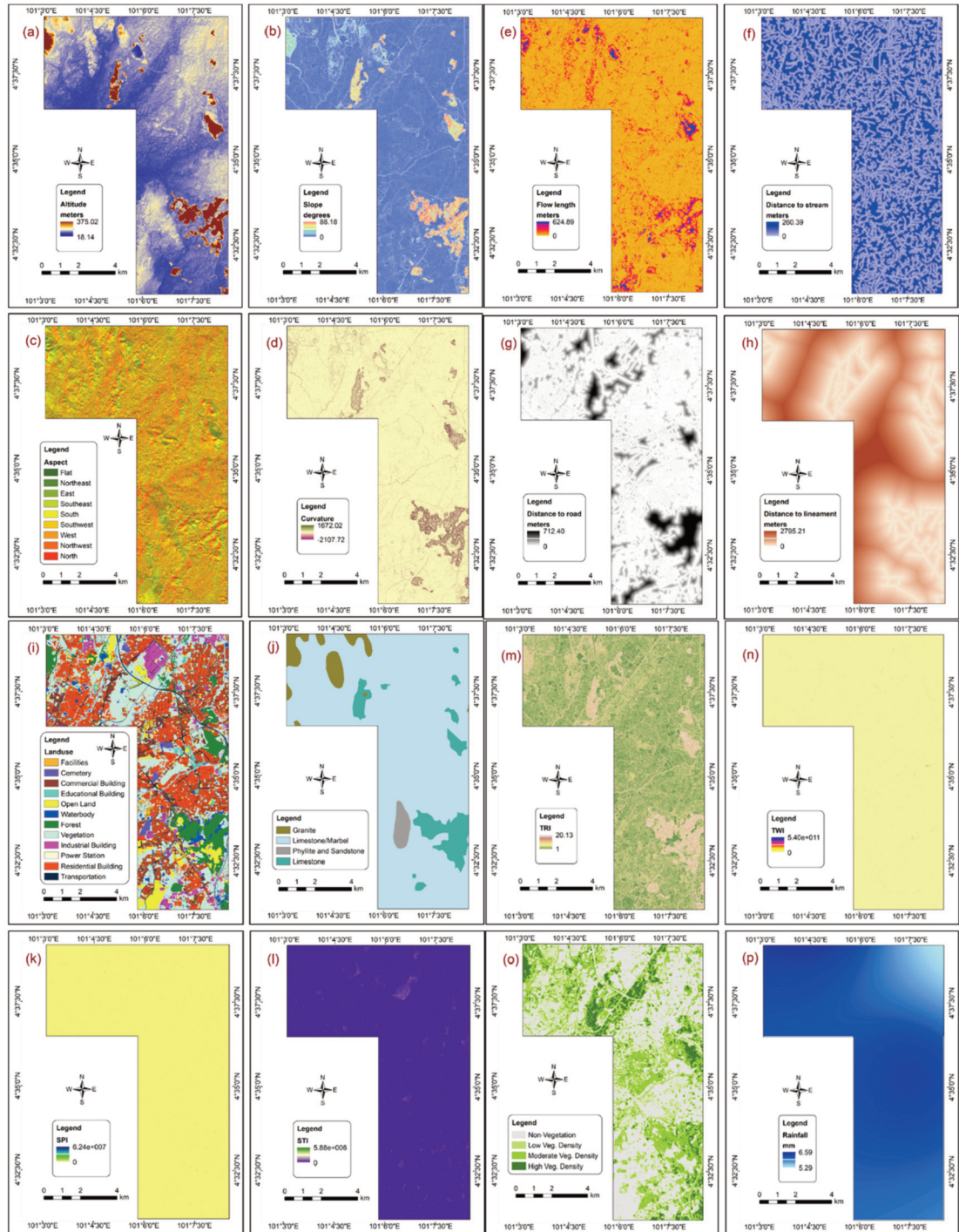


Fig. 2. Landslide conditioning factors.

stream power index (SPI), sediment transport index (STI), topographic roughness index (TRI) and topographic wetness index (TWI). The SPI is known as solid particle movement due to a combination of gravity that acts on sediments. Knowledge of sediment transport is commonly applied to define whether erosion or deposition will occur. The SPI is defined as follows (LeDell *et al.*, 2015):

$$SPI = A_s \times \tan\beta \quad (1)$$

where A_s and β are the catchment area and slope angle, respectively.

The STI describes the process of slope failure and deposition. This factor can be determined similarly to the following formula (LeDell *et al.*, 2015):

$$STI = \left(\frac{A_s}{22.13}\right)^{0.6} \times \left(\frac{\sin\beta}{0.0896}\right)^{1.3} \quad (2)$$

where A_s and β are the catchment area and slope angle, respectively.

The TRI is also an important factor that influences landslides, and it is calculated using the following formula (LeDell *et al.*, 2015):

$$TRI = \sqrt{\max^2 - \min^2} \quad (3)$$

where max and min are the highest and lowest cell values in the nine rectangular neighbourhoods of altitude, correspondingly, whereas TWI is a factor used to measure topographic controls on the hydrological process and defined on the basis of the slope and flow direction (LeDell *et al.*, 2015).

$$TWI = \ln\left(\frac{A_s}{\beta}\right) \quad (4)$$

where A_s and β are the catchment area and slope angle, respectively.

2) Overall Methodology

Fig. 3 illustrates the overall workflow of developing the proposed hybrid model. The workflow encompasses four main steps. The field and LiDAR datasets were obtained in the first step. Consequently, several landslide

and rockfall conditioning factors were derived. The inventory dataset in addition to the various conditioning factors was prepared. The second step was the pre-processing of the input dataset. The DTM of the study area was generated on the basis of the collected laser scanning dataset through the MCC and IDW methods. Noises and outliers were eliminated before generating the DTM. Furthermore, the georeferencing process was performed to convert the dataset from various sources into an identical format. In addition, the missing values of the inventory dataset were removed. The obtained conditioning factors were optimised using ant colony optimisation (ACO) and random forest (RF). These approaches were conducted to determine the optimal set of conditioning factors for landslide and rockfall.

The third step was the core-processing module in the proposed method, which consisted of developing the BANN and GMM. The probabilities of landslide and rockfall were produced through the developed BANN model which utilises the inventory dataset and conditioning factors. Several machine learning algorithms are existing and there is no agreement of which algorithm is the best because this relies on the application and the data samples. On the other hand, it is not applicable to evaluate all the existing algorithms. Therefore, this research employed three different algorithms that are commonly used in landslide studies and revealed a good accuracy. In addition, each algorithm has different characteristics. For instance, k NN is the simplest machine learning algorithm while ANN is more complex and requires big data samples. Whereas, SVM is effective in high dimensional spaces and it is also memory efficient. In addition, various kernel functions can be specified for different decision functions. A comprehensive comparative evaluation with other machine learning algorithms, such as SVM and k NN, and their ensembles was conducted, and then the BANN model was selected. The grid search method was adopted to select the hyperparameters of the machine learning algorithms. The standard

accuracy metrics, such as cross-validation area under the curve (CV-AUC), ROC curves and overall accuracy, were used to select the optimal model (Bruzzone and Prieto, 2000). Moreover, the GMM was developed to determine the distribution of slope angles and derive the prime geomorphological units of the study area. The GMM was trained using a slope dataset extracted from the generated DTM. The GMM hyperparameters were

determined through iterative search in accordance with the values of Bayesian information criterion (BIC) and Akaike information criterion (AIC). The outputs of the GMM were the thresholds of the slope angle (MUs) which allows an automatic detection of the probable source areas of various mass movement types.

Mapping, validations and model comparisons with other methods were the last step. The maps of landslide

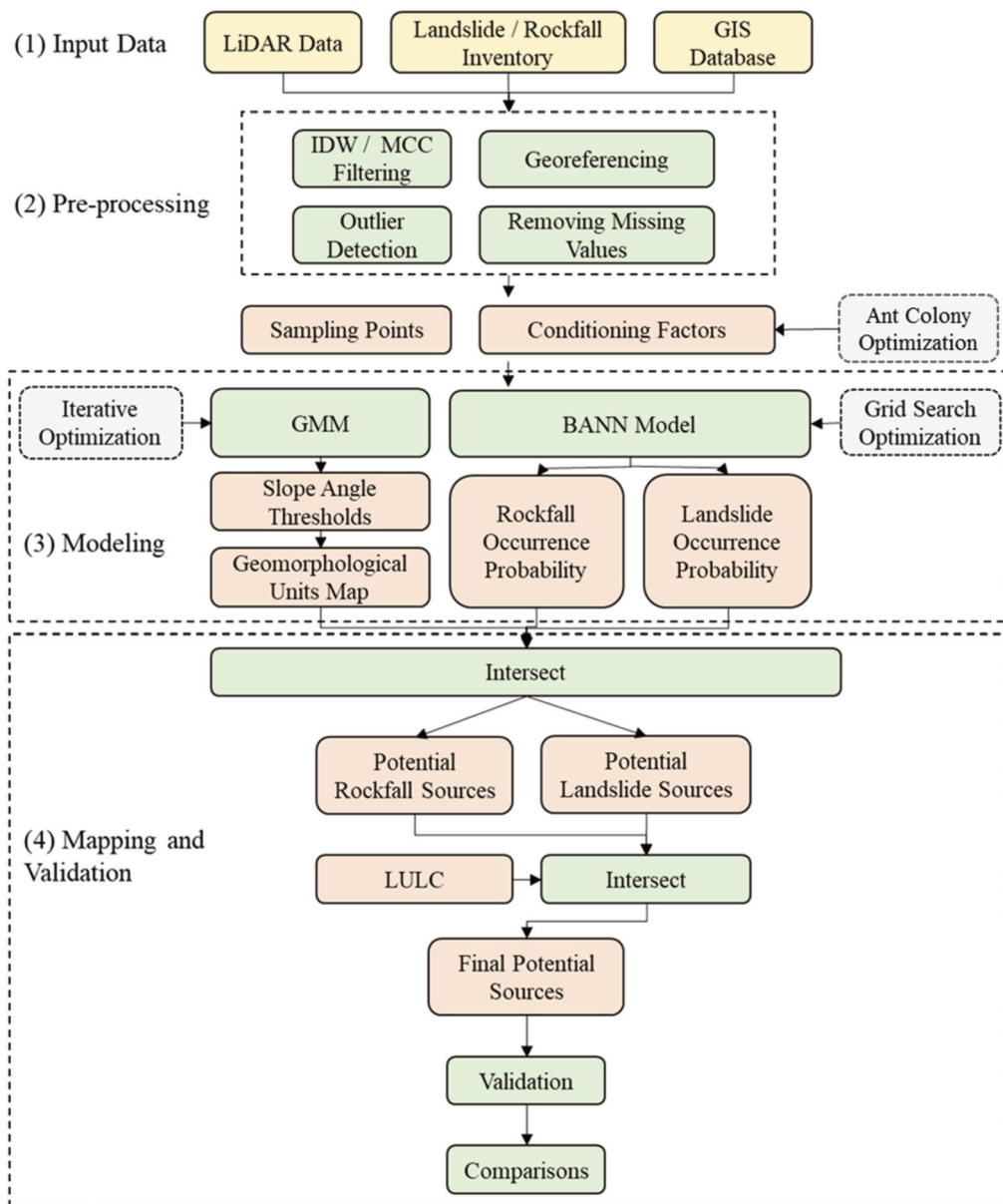


Fig. 3. Workflow of the proposed integrated model for detecting potential rockfall sources.

and rockfall probabilities obtained from the previous step were simultaneously utilised with a reclassified slope raster on the basis of the thresholds determined by the GMM to identify the probable source areas of landslide and rockfall. Then, the produced maps were intersected with the LULC map (only vegetated area, open land and forest were kept) to remove the noises and areas with a low probability of landslide and rockfall occurrence. Afterwards, the final produced maps of probable source areas of landslide and rockfall were validated on the basis of the testing inventory dataset and accuracy metrics, such as the ROC curves and confusion metric. In addition, a field investigation was conducted to validate the identified potential source areas.

(1) GMM

Assuming that the slope angle distribution can be modelled as the GMM, this technique performs an iterative evaluation of the GMM parameters and thus locates the slope angle thresholds. Expectation-maximisation (EM) algorithm can be used to determine model parameters from the data which iteratively modifies the GMM parameters that maximises the likelihood of the dataset (Skakun *et al.*, 2017). This algorithm has two main stages (i.e. expectation and maximisation). The expectation stage includes a fine enrolment of every observation to every component of the GMM. The maximisation stage offers a new parameter estimation. The expectation and maximisation stages were iterated until the model converges. Considering that the GMM components were specified, the Bayesian rule for the minimum error was applied to determine the optimal slope angle thresholds. The GMM is commonly utilised as a parametric model of the probability distributions expressed in the following formula (Skakun *et al.*, 2017):

$$p(x | \lambda) = \sum_{i=1}^k w_i g(x | \mu_i, \Sigma_i) \quad (5)$$

where x is the d-dimensional feature, $w_i, i=1, \dots, k$, are the mixture weights and $g(x | \mu_i, \Sigma_i), i=1, \dots, k$, are the component Gaussian densities. Each component density is a d-variate Gaussian function of the form (Skakun *et al.*, 2017):

$$g(x | \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i)\right\} \quad (6)$$

where μ_i is the mean vector, and Σ_i is the covariance matrix.

(2) ANN

The ANN is a computational model that is inspired by the human biological systems, such as process information and brain. An ANN model is formed by numerous strongly connected neurons. This model learns by example, which includes adjustments to the synaptic connections that exist between the neurons. A typical ANN model is frequently organised in layers. Every network layer is an array of neurons. Information flows through each neuron; each of them receives an input, processes it and forwards an output to the other linked neurons in the next layer. Multilayer perceptron (MLP) is a typical example of such a network (Fig. 4). This network normally has three layers of processing elements with only one hidden layer. However, no limitation is found on the hidden layer number. The input layer receives the external stimuli and propagates it to the adjacent layer. Furthermore, the mission of the hidden layer is to receive the weighted sum of incoming signals from the input units and processes it using an activation function. The saturation, hyperbolic tangent and sigmoid are the frequently utilised activation functions. The hidden units send an output signal towards the neurons in the adjacent layer. This next layer can be either the output or another hidden layer of arranged processing elements. The output layer units receive the weighted sum of incoming signals and process it using an activation function. Information is

forward propagated until the network produces an output (Humphrey *et al.*, 2017).

(3) Ensemble Modelling

Ensemble models are methods that combine multiple base models to create a highly robust one that can produce improved results. These models are frequently more accurate than the single ones (Youssef *et al.*, 2016; Pham *et al.*, 2017; Corsini and Mulas, 2017). Several ensemble approaches, such as bagging, voting and boosting, are available. Bagging (also known as bootstrap aggregating) is a standard ensemble learning method. The various classifiers in bagging were acquired through bootstrapped replication of the

training dataset. Thus, various subsets of training dataset were arbitrarily drawn, with replacement from the entire training dataset. Every subset was utilised to train various classifiers of the same type. Subsequently, single classifiers were integrated by taking a simple majority vote of their decisions. For any given example, the ensemble decision was the class selected through a highly classified number. Moreover, boosting creates multiple models of the same type, each of which learns to fix the prediction errors of a prior model in the chain. In addition, voting ensembles create multiple models (basically of different classifiers), and simple statistics, such as computing the mean and majority, are utilised to consolidate predictions.

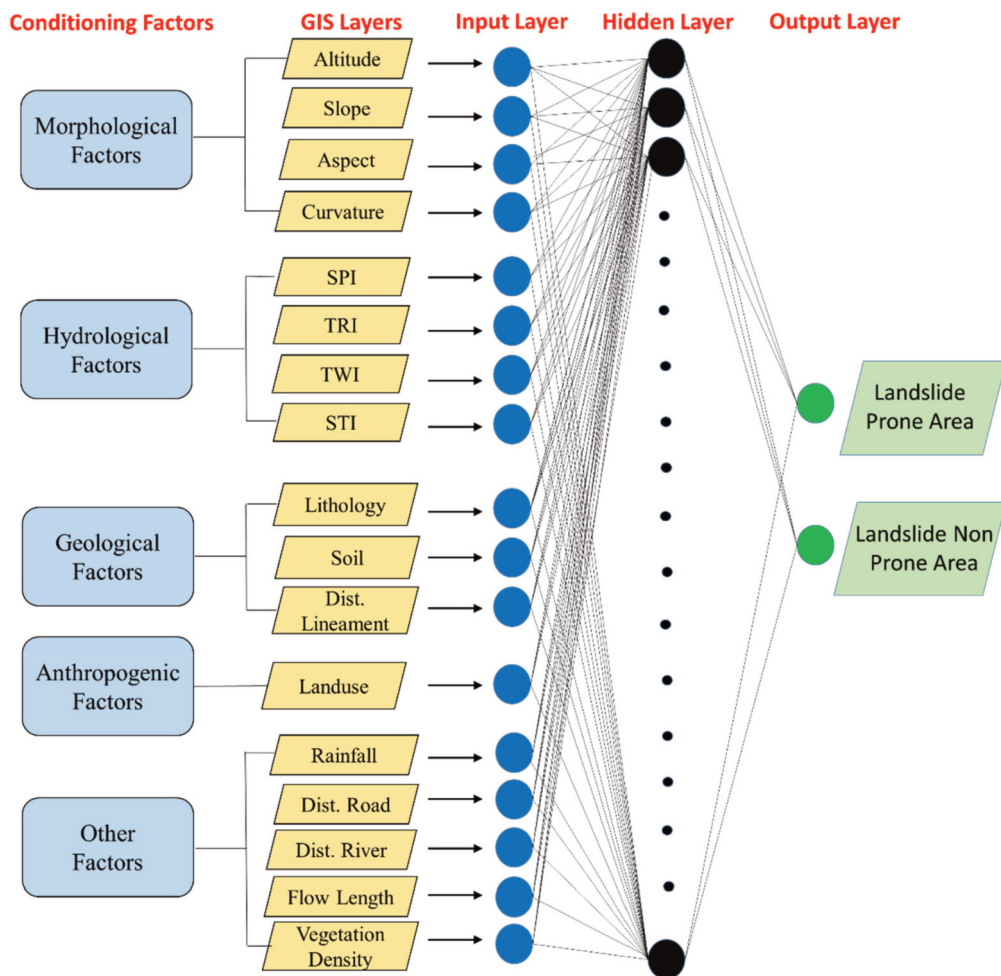


Fig. 4. Multi-layered feedforward neural network.

Table 1. Model hyperparameters that were optimised through the grid search method

Model	Parameters	Optimised Parameters	Search Space
<i>k</i> NN	<i>k</i> or number of neighbours	<i>k</i>	(1-10)
SVM	Kernel C Gamma (if C is not linear)	Kernel C Gamma	[Linear, RBF, Sigmoid, Polynomial] (1-100) by step of 1 [0, 1e-05, 0.0001, 0.001, 0.01, 0.1, 0.5, 0.9]
ANN	Batch size Solver Momentum Learning rate Activation Early stopping Hidden layer sizes	Batch size Solver Momentum Learning rate Activation Early stopping	[4,16] [LBFGS, SGD, Adam] [0, 0.3, 0.5, 0.9] [0.0001, 0.001, 0.01, 0.1, 0.9] [Logistic, Tanh, ReLU] [True, False]

In this research, several base models, such as ANN, *k*NN and SVM, were used. Three ensemble methods (i.e. bagging, boosting and voting) were investigated. The base models were optimised view grid search (presented in the next section). Then, the eminent model was decided in accordance with the CV-AUC. In bagging and boosting ensembles, the number of trees was 100. In voting ensembles, two models were combined, and a soft voting was used. The best fit ensemble model (BANN) was then used to produce the probabilities maps of shallow landslide and rockfall. This based on the inventory dataset of each landslide types in combination with the best subset of the conditioning factors obtained through ACO method. The model was run through Python environment and then the derived weights of each factor (relative importance) were used to produce the probability maps of each landslide types within GIS environment.

(4) Grid Search Optimisation of Base Models

A grid search is a standard search method for selecting sub-optimal hyperparameters of a machine learning/statistical model. Suppose that *k* parameters are present, and each of them has c_i values. Then, the number of search possibilities (P) is

$$P = \prod_{k=1}^{i=1} c_i \quad (7)$$

Table 1 lists the hyperparameters of the base models (ANN, *k*NN and SVM) along with their optimised parameters. For example, the SVM model has three parameters (*k*), and the kernel function has four values (i.e. linear, RBF, sigmoid and polynomial, C), which have 100 values from 1 to 100 and gamma that exists only for non-linear kernels. Thus, the number of search possibilities (P) is $3 \times 100 \times 8 + 1 \times 100 = 2500$. These possibilities were tested through the CV method, and the optimal combination of parameters was decided on the basis of the prediction accuracies obtained.

(5) Accuracy Metrics

The success (ROC) and prediction (PRC) curves were utilised to validate the proposed hybrid model for identifying the probable landslide and rockfall in Ipoh. The ROC and PRC curves demonstrate the known rockfall percentage that lay on probability level ranks and show the graph of the cumulative frequency (Dou *et al.*, 2018; Jeong *et al.*, 2018; Pham *et al.*, 2016). The success and prediction rates were generated on the basis of the training and validation data subset of rockfalls, respectively. Furthermore, the AUC can be utilised to define the accuracy of the probability maps qualitatively, in which a large AUC means a high accuracy achieved (Samia *et al.*, 2017; Yan *et al.*, 2018; Hong *et al.*, 2017).

4. Results and Discussions

The major findings obtained from this research are presented in this section. Firstly, a summary statistics of modelling data and its pre-processing are provided. Secondly, the results of the GMM are presented. Subsequently, the results of the BANN model, including the probability maps, optimisation results and comparisons with other methods (*k*NN and SVM), are explained. Finally, the validation and field verification are discussed.

1) Summary Statistics and Pre-processing

The inventory data had 147 sampling points (83 rockfalls and 64 landslides). The slope angles in the landslide data samples ranged from 19° to 49° with an average slope angle of 34.75° (std. = 13.43°). By contrast, the slope angles in the rockfall data samples had a minimum of 47° and a maximum of 76.39°. The average slope angle was 65.86°, and the standard deviation was 11.34°.

In addition, the multicollinearity of the factors was analysed through variance inflated factor (VIF) method because the sampling points (landslide and rockfall) are subjected to strong correlations in different conditioning factors. Table 2 lists the VIF values calculated among the conditioning factors in the landslide and rockfall dataset samples. According to a previous study Hong *et al.*, (2017), a VIF value of greater than 4 is considered highly collinear. Thus, the corresponding factors should be removed from further analysis. The highest VIF values were 3.107 and 3.272 for slope and STI factors in the landslide and rockfall data samples, correspondingly. Consequently, none of the factors was removed.

Moreover, the entire conditioning factors were optimised using ACO to determine the optimal subsets of conditioning factors for identifying the potential areas of landslide and rockfall occurrences accurately. The assessment of the optimal subset of the conditioning factors was performed on the basis of the RF algorithm. The ACO revealed that the probability of rockfall

Table 2. VIF values calculated among the conditioning factors in the landslide and rockfall data samples

Factor	Landslide		Rockfall	
	Multiple R ²	VIF	Multiple R ²	VIF
Vegetation density	0.417	1.716	0.464	1.867
Soil	0.375	1.600	0.331	1.495
Land use	0.454	1.832	0.478	1.917
Geology	0.455	1.833	0.410	1.694
Rainfall	0.308	1.446	0.268	1.367
Aspect	0.290	1.408	0.331	1.494
Curvature	0.264	1.359	0.201	1.252
Altitude	0.662	2.960	0.671	3.042
Flow length	0.412	1.702	0.449	1.816
Distance to road	0.491	1.963	0.416	1.712
Distance to lineament	0.605	2.535	0.650	2.860
Slope	0.678	3.107	0.685	3.178
SPI	0.547	2.209	0.623	2.650
STI	0.610	2.564	0.694	3.272
Distance to stream	0.304	1.437	0.330	1.492
TRI	0.379	1.610	0.430	1.753
TWI	0.302	1.432	0.333	1.499

occurrences can be determined on the basis of 9 conditioning factors, that is, slope, altitude, TRI, distances to lineaments, rivers and roads, geology, rainfall and vegetation density, with an accuracy of 83%. For shallow landslide, the optimal subset of the conditioning factors for identifying the potential areas of occurrences (with an accuracy of 80%) encompasses 13 of the conditioning factors, namely, slope, aspect, curvature, TWI, STI, TRI, geology, distances to lineaments, rivers and roads, vegetation density, rainfall and land use.

2) Results of the GMM–Slope Angle Distribution (SAD)

Table 3 displays the SADs as calculated through the optimised GMM. The optimal k value was determined to be 5 for all the three sub-study areas. The average number of iterations and regularisation value were 500 and 0.01, respectively. The means and standard deviations of five SAD (MU) were demonstrated for the three areas in Table 3. For the Gunung Lang area, the GMM determined 5 MU values (i.e. 1.78, 6.07, 16.04, 41.25 and 63.54) for five geomorphological units (i.e. plains, foot slopes, moderately steep slopes,

steep slopes and cliffs). The plain unit indicated low-slope angles that correspond to the fluvial and fluvio-glacial deposits. Foot slope is a gentle slope angle that features the lower part of the hillslope characterised by colluvial fans, debris flow and landslide deposits. Moderately steep and steep slopes are the units that contain deposits and rocky outcrops covered with vegetation. Furthermore, cliffs are very steep slopes which correspond to rocky outcrops. The SAD values for the Gua Tambun area were 1.92, 5.18, 13.35, 37.60 and 63.78. By contrast, the MU values for the Gunung Rapat area were 1.46, 6.23, 16.43, 43.21 and 66.31.

The SADs in the three study areas and Gaussian distribution were plotted on the basis of the mean (MU) and standard deviation of the SAD derived from the GMM (Fig. 5). The thresholds of the slope angle were specified by intersecting the Gaussian curves of various geomorphological units. The SADs of the three study areas could be produced accurately by the sum of Gaussian MU (GDMU) with a coefficient of determination of approximately 1. In particular, the morphologies could be described by the SAD decomposition. The intersecting cliff with steep slopes was modelled at 57°. According to literature, slope

Table 3. SADs determined by the GMM and optimal k values

Study Area	Optimum k	MU	Variance	Std.
Gunung Lang	5	1.78	0.567	0.753
		6.07	6.021	2.454
		16.04	36.044	6.004
		41.25	94.140	9.703
		63.54	78.246	8.846
Gua Tambun	5	1.92	0.593	0.770
		5.18	3.866	1.966
		13.35	27.647	5.257
		37.60	160.886	12.684
		63.78	95.162	9.755
Gunung Rapat	5	1.46	0.541	0.736
		16.43	38.071	6.170
		6.23	6.619	2.573
		43.21	126.370	11.241
		66.31	70.502	8.397

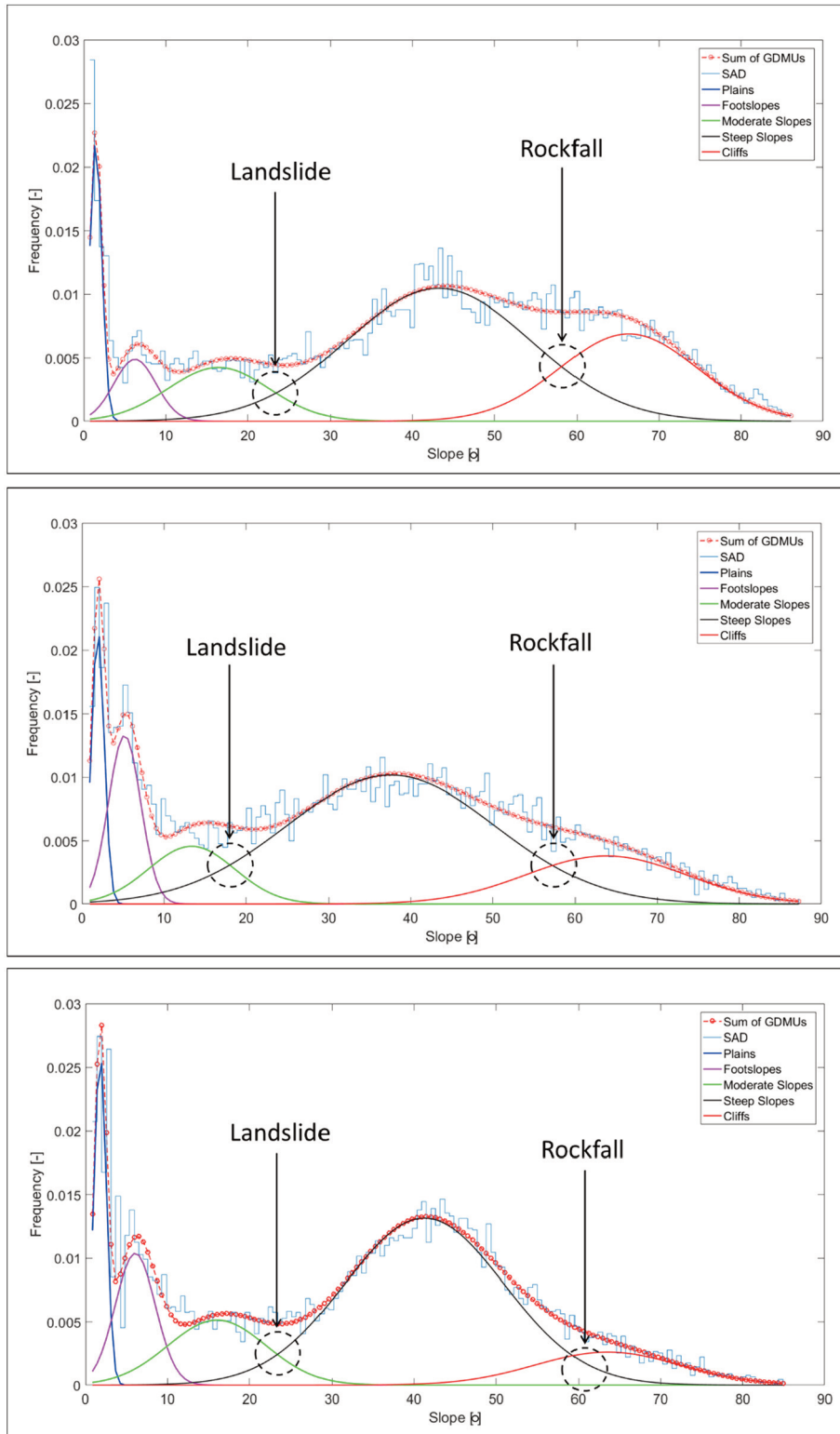


Fig. 5. Estimated distribution of slope angle: (a) Gunung Rapat, (b) Gua Tambun and (c) Gunung Lang areas.

thresholds of $>60^\circ$ and $>45^\circ$ were proposed by Guzzetti *et al.* (2003) and Jaboyedoff and Labiouse (2003) for detecting potential rockfall sources. The obtained rockfall slope angle threshold through the GMM was similar to that of Guzzetti *et al.* (2003). The small thresholds proposed by Jaboyedoff and Labiouse (2003) could be inefficient for Ipoh and could lead to misclassification of rockfall with other landslide types

because they have similar thresholds of slope angle. Thus, selecting a high slope angle as a threshold for the study area (i.e. Ipoh) is necessary to prevent any confusion among various types of unpredictable landslides and rockfalls.

On the basis of the determined slope angle thresholds, the slope raster of the study areas was reclassified into different geomorphological units. In Gunung Lang, plain and steep slopes were dominant. The landslide and rockfall inventories on this map showed that most landslides have occurred in steep slopes and rockfalls in cliffs (Fig. 6(a)). Similarly, in the Gua Tambun area, the dominant geomorphological units are plain and steep slopes (Fig. 6(b)). By contrast, in the Gunung Rapat area, the dominant units are plain, steep slopes and cliffs. The northeast part of the area is hilly and rugged. The landslide and rockfall inventories are placed in steep slope and cliff areas (Fig. 6(c)).

3) Results of Optimisation

Table 4 summarises the results of grid search optimisation on the base model's hyperparameters. Different parameter values were selected for landslide and rockfall data samples. The optimal k parameter values were 7 and 5 for landslide and rockfall models, respectively. For the SVM model, the linear kernel

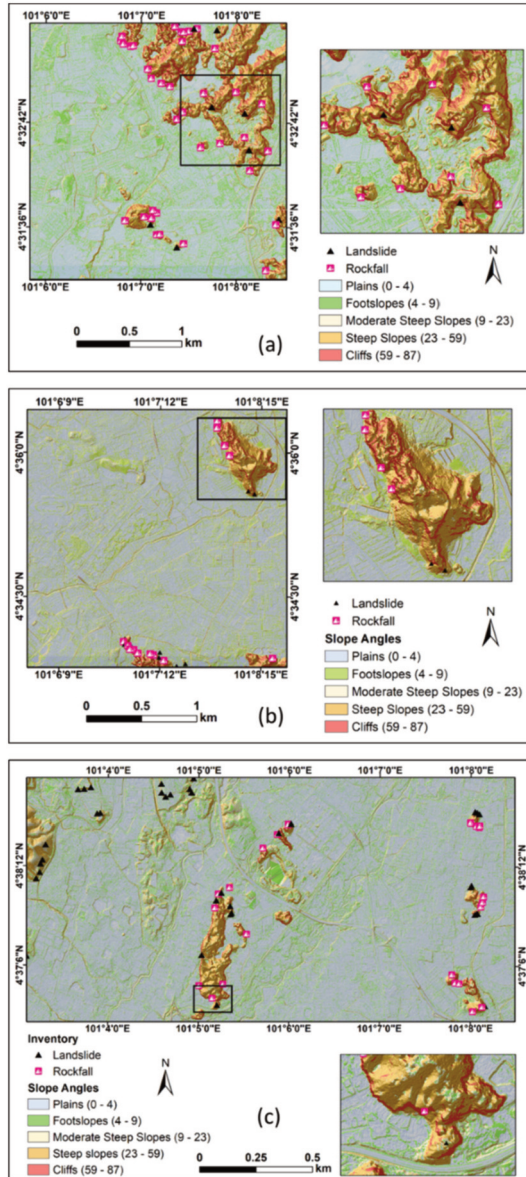


Fig. 6. Identified geomorphological units in (a) Gunung Rapat, Gua Tambun (b) and (c) Gunung Lang areas.

Table 4. Results of the grid search optimisation of the base model's hyperparameters

Model	Optimised Parameters	Best Value	
		Landslide	Rockfall
KNN	k	7	5
SVM	Kernel	Linear	Linear
	C	10	91
	Gamma		
ANN	Batch size	4	4
	Solver	LBFGS	LBFGS
	Momentum	0.9	0
	Learning rate	0.01	0.0001
	Activation	Tanh	Tanh
	Early stopping	True	True

function was highly suitable for both datasets. However, different C values (i.e. 10 and 91) were optimal for landslide and rockfall data samples. In terms of the ANN model, the grid search algorithm found that a batch size of four, LBFGS optimisation solver and 'Tanh' activation function are more suitable than the other values explored for both datasets. In addition, the optimisation process indicated that early stopping is essential for model generalisation. However, the learning rates of 0.01 and 0.0001 were optimal for landslide and rockfall data samples, correspondingly. The ensemble models were developed on the basis of the optimised base models listed in Table 4.

4) Results of the BANN and Source Identification

The produced probability maps of landslide and rockfall based on the BANN model are depicted in Fig. 7. The proposed ensemble model estimated the

occurrence probability, which ranged from 1 (high potential of landslide/rockfall occurrence) to 0 (no potential of landslide/rockfall occurrence). Nevertheless, the probability maps were reclassified into five classes using a quantile method to facilitate map interpretation. The classes were very high, high, moderate, low and very low. The final probability maps illustrate that Gunung Rapat, Gua Tambun and Gunung Lang areas have high probabilities of encountering landslide and rockfall events.

The intersections of slope angle thresholds obtained through the GMM and probability maps resulted in the potential landslide and rockfall source areas. This section presents the results of detecting potential landslide and rockfall source areas. The potential sources of landslides were detected by intersecting the slope angle threshold (23°) and probability values estimated by the BANN model. Fig. 8 demonstrates the results of potential landslide sources in the study areas. The maps show the BANN probabilities, wherein the

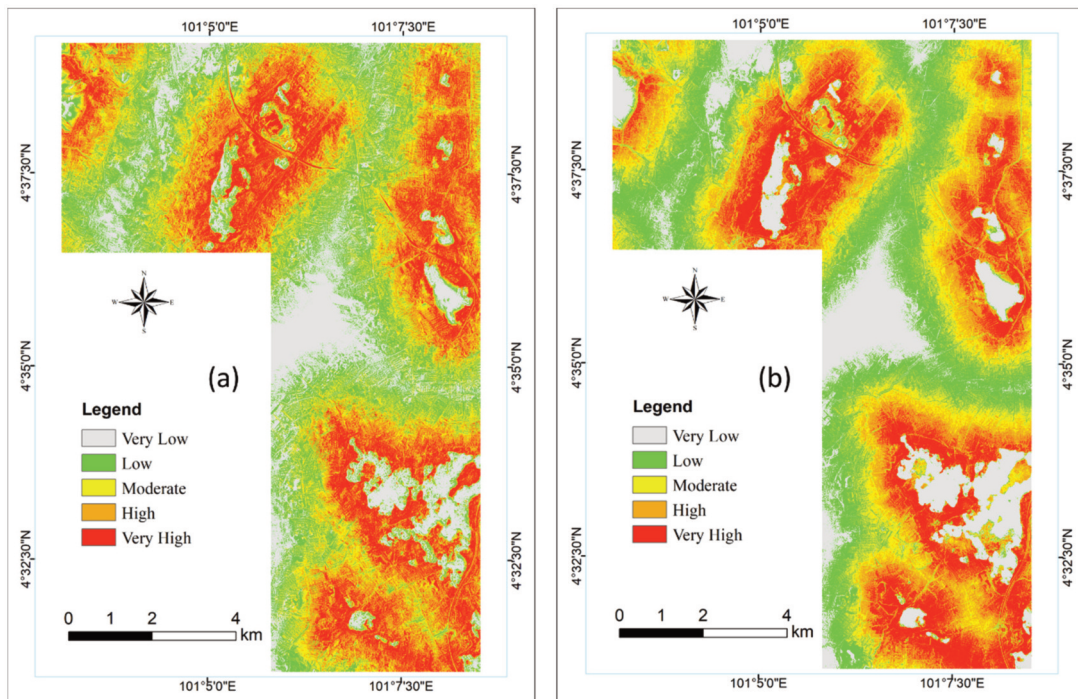


Fig. 7. Probability maps of (a) landslide and (b) rockfall occurrences.

slope angles are above the selected threshold. The areas with a slope less than the selected threshold were considered (highlighted by light grey). The landslide inventories are generally situated in regions detected as probable source areas, thereby indicating the robustness of the model. The quantitative assessments will be presented later.

The probable rockfall sources were determined by crossing the rockfall probability map produced by the

BANN model and obtained threshold of slope angle (57°) through the GMM. Fig. 9 exhibits the results of potential rockfall sources in the study areas. The rockfall inventories are mainly situated in regions detected as probable source areas. This result indicates that the proposed hybrid model can efficiently identify rockfall and landslide source areas in the study area. The decision maker can take into account the obtained results in the designing and development processes to

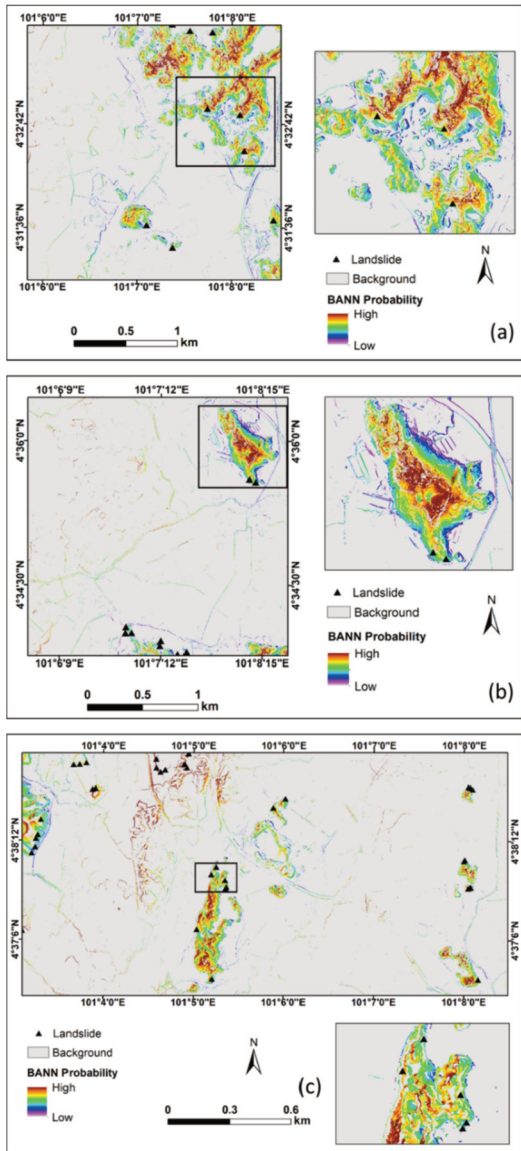


Fig. 8. Identified potential sources of landslide in (a) Gunung Rapat, (b) Gua Tambun and (c) Gunung Lang areas.

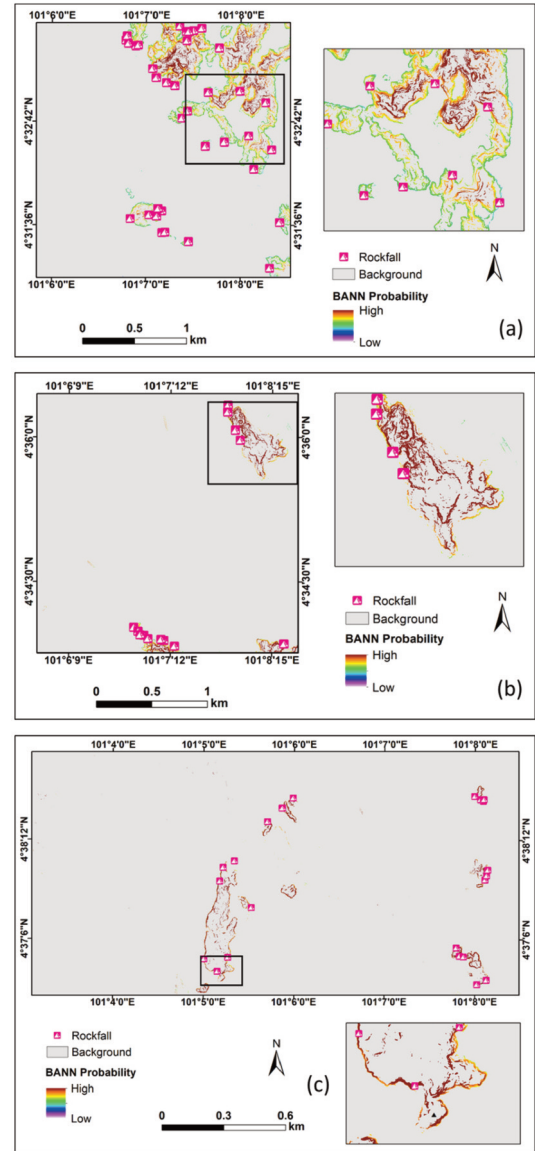


Fig. 9. Identified potential sources of rockfall in (a) Gunung Rapat, (b) Gua Tambun and (c) Gunung Lang areas.

Table 5. Comparisons of the BANN with other ensemble methods

Dataset	Accuracy Metric	kNN	SVM	ANN	BkNN	BSVM	BANN	AdaSVM	VkNN-SVM	VkNN-ANN	VSVM-ANN
Landslide	Training accuracy	0.733	0.567	0.789	0.756	0.644	0.852	0.522	0.744	0.789	0.800
	Testing accuracy	0.667	0.615	0.564	0.692	0.487	0.821	0.513	0.744	0.692	0.667
	Training AUC	0.864	0.708	0.820	0.850	0.687	0.896	0.497	0.808	0.887	0.883
	Testing AUC	0.729	0.725	0.688	0.743	0.529	0.854	0.545	0.749	0.804	0.701
	CV-AUC	0.736	0.633	0.732	0.700	0.632	0.836	0.542	0.677	0.751	0.672
Rockfall	Training accuracy	0.852	0.755	0.890	0.841	0.733	0.950	0.604	0.798	0.900	0.910
	Testing accuracy	0.897	0.872	0.871	0.872	0.848	0.926	0.604	0.897	0.901	0.897
	Training AUC	0.925	0.905	0.910	0.939	0.901	0.950	0.959	0.907	0.910	0.910
	Testing AUC	0.940	0.903	0.908	0.944	0.910	0.948	0.965	0.919	0.925	0.918
	CV-AUC	0.864	0.870	0.895	0.834	0.870	0.955	0.931	0.892	0.911	0.901

protect people from the hazard of such incidents and have a sustainable environment. Moreover, as the rockfall source areas are the key element in the assessment of rockfall hazard and risk, the derived results by this research can be used for the modelling of rockfall trajectories and their characteristics.

5) Model Comparison

The results of the proposed ensemble BANN model were validated by comparing this output with the base and other ensemble models, such as voting, boosting and bagging, using the same base models as summarised in Table 4. The proposed BANN model was compared with 10 other models. Table 5 presents the findings of the comparative experiments and provides various model accuracy metrics for the data samples of landslides and rockfalls. In general, the ensemble models involving voting and bagging outperform the boosting and individual models on most accuracy measures and both datasets. Nevertheless, the boosting model based on the SVM revealed a low performance on both datasets. The proposed BANN model achieved the highest accuracy and optimal outputs for data samples of landslides and rockfalls among the models. The rockfall data achieved the optimal accuracies in all the accuracy matrices over other methods. The BANN achieved 95% of training

accuracy and 0.955 of fivefold CV-AUC. In addition, it achieved 93% of testing accuracy and 0.950 of testing AUC. For the landslide data, BANN achieved the optimal training accuracy of 85.2% and 0.836 of fivefold CV-AUC. It also achieved 82% of testing accuracy and 0.854 of testing AUC. Thereby indicating that the model can be generalised and replicated in different regions, and the proposed method can be applied to various landslide studies. Among the base models, kNN obtained the optimal outputs in all the accuracy metrics, except the training accuracy on the landslide dataset. For rockfall data, the ANN accomplished the optimal outputs in all the accuracy metrics. By contrast, the SVM was poorer on both datasets than the kNN and ANN algorithms.

6) Research Limitation

The proposed methods were applied and the main goal of identifying rockfall source areas in presence of other landslide types was achieved. However, some limitations should be considered in future research. For instance, the temporal factor was not considered thus the return period assessment was not performed. This is because the inventory dataset is not complete in time space. The geomechanical characteristics such as fractures and discontinuities were not taken into account. However, such information requires extensive

field surveys which are costly and time consuming and it is difficult to perform field survey for regional scale study. In addition, the focus of this study to assess the LiDAR dataset as alternative of conducting in-situ investigations. Moreover, it is hard to obtain realistic results using the proposed methods where the inventory dataset is not available or limited.

7) Field Verification

Several field observations were performed to verify the outcomes of the rockfall source modelling. Various locations that cover the entire study area were randomly selected to conduct site verification. The in-situ survey was conducted using a GNSS technique to identify the

predicted source locations and compare these locations with the real field condition. In addition, geomechanical survey was performed to assess the presence of discontinuities and fractures. According to the field observations, all locations were found within the probable rockfall source areas (Fig. 10), which correspond to the rockfall source modelling. In addition, discontinuities and fractures evidently appeared in all locations. According to the interviews that have been conducted in different locations, several rockfall incidents were triggered by a large group of monkeys that live in these locations in addition to climate factors (rainfall and wind).

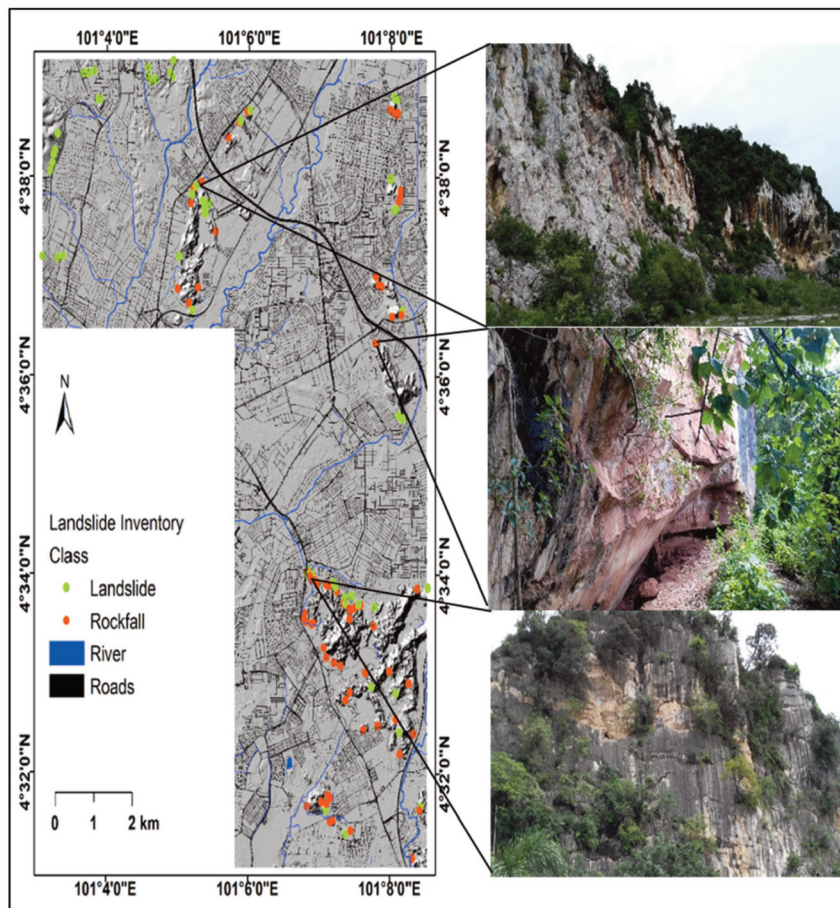


Fig. 10. Field verification.

5. Conclusion

This research proposed a hybrid model by combining two approaches (BANN and GMM) for identifying probable rockfall sources at the Ipoh area using the LiDAR dataset. The major motivation behind this research was the rareness of techniques in the literature, which can detect rockfall source areas accurately with the presence of other types of landslides, such as a shallow one. Thus, the current research successfully handled this problem.

Various machine learning algorithms (k NN, SVM, and ANN) were tested individually and with different ensemble models (boosting, bagging, and voting) to identify the best fit model that can accurately produce the probabilities of landslide and rockfall. The ensemble BANN model was combined with the GMM in a monocular framework for detecting the high probability of landslide/rockfall occurrences and solving the automatic determination of slope thresholds. The BANN model accomplished the optimal results for landslide and rockfall probabilities among the investigated models, such as voting and bagging models. This model achieved 0.836 and 0.955 of fivefold CA-AUC for the landslide and rockfall modelling, correspondingly. Moreover, the GMM accurately determined the distributions and identified the thresholds of the slope angle that allowed the source identification of landslide and rockfall at 23° and 59°, respectively. LiDAR technique proved to be an efficient alternative of geomechanical survey which is costly and time consuming.

This research also showed the necessity of optimising the base models to achieve the optimal possible results. For example, the optimal k values of the k NN model were 7 and 5 for landslide and rockfall datasets, correspondingly. In addition, the optimisation results suggested that using early stopping in ANN models is crucial to achieving excellent generalisation capabilities and preventing overfitting. However, the search for

all the combinations of the hyperparameters was computationally expensive. Thus, using robust methods to fine-tune the models is suggested. Other areas of improvements of this research include standardising the model for multi-tasking. In the current research, different optimal parameters were found for landslide and rockfall modelling. Generally, this research contributes to the lack in literature of identifying rockfall source areas in presence other landslide types in a semi-automatic way. The proposed methods can be applied in different areas and expected to perform well. This because of the slope threshold can be identified automatically based on the geomorphological units and the ensemble model reviled a good accuracy based on both training and testing dataset. The novelty of the current research is proposing a hybrid model based on an ensemble model that never tested in rockfall studies and an automatic method for the determining slope thresholds. The obtained results can highly assist in designing the development of urban area and prevent people from encountering landslide and rockfall hazardous. In addition, the identified source areas can be used for further assessment of rockfall hazard, vulnerability, and risk. However, finding a single model that can precisely detect the landslide and rockfall occurrences with the same architecture and parameters is important.

Acknowledgements

The authors wish to thank the Department of Mineral and Geosciences, the Department of Surveying Malaysia, the Federal Department of Town and Country Planning Malaysia for the data provided. This research was supported by the UTS under grant number 321740.2232335 and 321740.2232357 and was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2018M1A3A3A02066008).

References

- Acosta, E., F. Agliardi, G.B. Crosta, and S. Ríos Aragues, 2007. Regional rockfall hazard assessment in the Benasque Valley (Central Pyrenees) using a 3D numerical approach, *Proc. of 4th EGS Plinius Conference Mediterranean Storms*, Mallorca, Oct. 2-4, pp. 555-563.
- Agliardi, F., F. Riva, L. Galletti, A. Zanchi, and G.B. Crosta, 2016. Rockfall source characterization at high rock walls in complex geological settings by photogrammetry, structural analysis and DFN techniques, *EGU General Assembly Conference Abstracts*, Vienna, Apr. 17-22, vol. 18, p. 1307.
- Bruzzzone, L. and D.F. Prieto, 2000. Automatic analysis of the difference image for unsupervised change detection, *IEEE Transactions on Geoscience and Remote Sensing*, 38(3): 1171-1182.
- Budetta, P., 2004. Assessment of rockfall risk along roads, *Natural Hazards and Earth System Science*, 4(1): 71-81.
- Budetta, P., C. De Luca, and M. Nappi, 2016. Quantitative rockfall risk assessment for an important road by means of the rockfall risk management (RO. MA.) method, *Bulletin of Engineering Geology and the Environment*, 75(4): 1377-1397.
- Corona, C., J. Lopez-Saez, A. Favillier, R. Mainieri, N. Eckert, D. Trappmann, M. Stoffel, F. Bourrier, and F. Berger, 2017. Modeling rockfall frequency and bounce height from three-dimensional simulation process models and growth disturbances in submontane broadleaved trees, *Geomorphology*, 281: 66-77.
- Corona, C., D. Trappmann, and M. Stoffel, 2013. Parameterization of rockfall source areas and magnitudes with ecological recorders: when disturbances in trees serve the calibration and validation of simulation runs, *Geomorphology*, 202: 33-42.
- Corsini, A. and M. Mulas, 2017. Use of ROC curves for early warning of landslide displacement rates in response to precipitation (Piagneto landslide, Northern Apennines, Italy), *Landslides*, 14(3): 1241-1252.
- Dou, J., H. Yamagishi, Z. Zhu, A.P. Yunus, and C.W. Chen, 2018. TXT-tool 1.081-6.1 A Comparative Study of the Binary Logistic Regression (BLR) and Artificial Neural Network (ANN) Models for GIS-Based Spatial Predicting Landslides at a Regional Scale, In: Kyoji, S., Guzzetti, F., Yamagishi, H., Arbanas, Ž., Casagli, N., McSaveney, M., Dang, K. (Eds.), *Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools*, Springer, Berlin, Germany, vol. 1, pp. 139-151.
- Evans, J.S. and A.T. Hudak, 2007. A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments, *IEEE Transactions on Geoscience and Remote Sensing*, 45(4): 1029-1038.
- Fanos A.M., B. Pradhan, S. Mansor, Z.M. Yusoff, and A.F. bin Abdullah, 2018. A hybrid model using machine learning methods and GIS for potential rockfall source identification from airborne laser scanning data, *Landslides*, 15(9): 1833-1850.
- Fanos, A.M. and B. Pradhan, 2018. Laser scanning systems and techniques in rockfall source identification and risk assessment: a critical review, *Earth Systems and Environment*, 1-20.
- Fanos, A.M. and B. Pradhan, 2016. Multi-scenario Rockfall Hazard Assessment Using LiDAR Data and GIS, *Geotechnical and Geological Engineering*, 34(5): 1375-1393.
- Fanos, A.M., B. Pradhan, A.A. Aziz, M.N. Jebur, and H.J. Park, 2016. Assessment of multi-scenario rockfall hazard based on mechanical parameters using high-resolution airborne laser scanning

- data and GIS in a tropical area, *Environmental Earth Sciences*, 75(15): 1129.
- Gigli, G., S. Morelli, S. Fornera, and N. Casagli, 2014. Terrestrial laser scanner and geomechanical surveys for the rapid evaluation of rock fall susceptibility scenarios, *Landslides*, 11(1): 1-14.
- Guzzetti, F., P. Reichenbach, and G.F. Wieczorek, 2003. Rockfall hazard and risk assessment in the Yosemite Valley, California, USA, *Natural Hazards and Earth System Science*, 3(6): 491-503.
- Hong, H., B. Pradhan, M.I. Sameen, W. Chen, and C. Xu, 2017. Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China), *Natural Hazards and Risk*, 8(2): 1-26.
- Humphrey, G.B., H.R. Maier, W. Wu, N.J. Mount, G.C. Dandy, R.J. Abrahart, and C.W. Dawson, 2017. Improved validation framework and R-package for artificial neural network models, *Environmental Modelling & Software*, 92: 82-106.
- Jaboyedoff, M. and V. Labiouse, 2003. Preliminary assessment of rockfall hazard based on GIS data, *Proc. of 10th ISRM Congress*, Sandton, South Africa, Sep. 8-12.
- Jeong, S., A. Kassim, M. Hong, and N. Saadatkhah, 2018. Susceptibility Assessments of Landslides in Hulu Kelang Area Using a Geographic Information System-Based Prediction Model, *Sustainability*, 10(8): 2941.
- LeDell, E., M. Petersen, and M. Van der Laan, 2015. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates, *Electronic Journal of Statistics*, 9(1): 1573.
- Loye, A., M. Jaboyedoff, and A. Pedrazzini, 2009. Identification of potential rockfall source areas at a regional scale using a DEM-based geomorphometric analysis, *Natural Hazards and Earth System Sciences*, 9(5): 1643-1653.
- Malamud, B.D., D.L. Turcotte, F. Guzzetti, and P. Reichenbach, 2004. Landslide inventories and their statistical properties, *Earth Surface Processes and Landforms*, 29(6): 687-711.
- Messenzehl, K., H. Meyer, J.C. Otto, T. Hoffmann, and R. Dikau, 2017. Regional-scale controls on the spatial activity of rockfalls (Turtmann valley, Swiss Alps) a multivariate modeling approach, *Geomorphology*, 287: 29-45.
- Mineo, S., G. Pappalardo, M. Mangiameli, S. Campolo, and G. Mussumeci, 2018. Rockfall Analysis for Preliminary Hazard Assessment of the Cliff of Taormina Saracen Castle (Sicily), *Sustainability*, 10(2): 417.
- Moos, C., M. Fehlmann, D. Trappmann, M. Stoffel, and L. Dorren, 2018. Integrating the mitigating effect of forests into quantitative rockfall risk analysis-two case studies in Switzerland, *International Journal of Disaster Risk Reduction*, 32: 55-74.
- Mote, T.I., M.D. Skinner, M.L. Taylor, and C. Lyons, 2019. Site-Specific Rockfall Risk Assessments and Rockfall Protection Structure Design Following the 2010/2011 Canterbury Earthquake Sequence, *Proc. of IAEG/AEG Annual Meeting*, San Francisco, CA, vol. 5, pp. 143-152.
- Muzzillo, R., L. Losasso, and F. Sdao, 2018. Rockfall Source Areas Assessment in an Area of the Pollino National Park (Southern Italy), *Proc. of International Conference on Computational Science and Its Applications ICCSA 2018*, Melbourne, VIC, Jul. 2-5, vol. 10962, pp. 366-379.
- Pellicani, R., G. Spilotro, and C.J. Van Westen, 2016. Rockfall trajectory modeling combined with heuristic analysis for assessing the rockfall hazard along the Maratea SS18 coastal road (Basilicata, Southern Italy), *Landslides*, 13(5): 1143-1153.

- 985-1003.
- Pham, B.T., D.T. Bui, I. Prakash, and M.B. Dholakia, 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS, *Catena*, 149: 52-63.
- Pham, B.T., B. Pradhan, D.T. Bui, I. Prakash, and M.B. Dholakia, 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India), *Environmental Modelling & Software*, 84: 240-250.
- Pourghasemi, H., A. Gayen, S. Park, C.W. Lee, and S. Lee, 2018. Assessment of Landslide-Prone Areas and Their Zonation Using Logistic Regression, LogitBoost, and NaïveBayes Machine-Learning Algorithms, *Sustainability*, 10(10): 3697.
- Pradhan, B., M.H. Abokharima, M.N. Jebur, and M.S. Tehrani, 2014. Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS, *Natural Hazards*, 73(2): 1019-1042.
- Pradhan, B. and A.M. Fanos, 2017a. Application of LiDAR in Rockfall Hazard Assessment in Tropical Region, In: Pradhan, B. (Eds.), *Laser Scanning Applications in Landslide Assessment*, Springer, Berlin, Germany, pp. 323-359.
- Pradhan, B. and A.M. Fanos, 2017b. Rockfall hazard assessment: an overview, In: Pradhan, B. (Eds.), *Laser Scanning Applications in Landslide Assessment*, Springer, Berlin, Germany, pp. 299-322.
- Samia, J., A. Temme, A. Bregt, J. Wallinga, F. Guzzetti, F. Ardizzone, and M. Rossi, 2017. Characterization and quantification of path dependency in landslide susceptibility, *Geomorphology*, 292: 16-24.
- Skakun, S., B. Franch, E. Vermote, J.C. Roger, I. Becker-Reshef, C. Justice, and N. Kussul, 2017. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model, *Remote Sensing of Environment*, 195: 244-257.
- Varnes, D.J., 1978. Slope movement types and processes, *Special Report*, 176: 11-33.
- Yan, G., S. Liang, X. Gui, Y. Xie, and H. Zhao, 2018. Optimizing landslide susceptibility mapping in the Kongtong District, NW China: comparing the subdivision criteria of factors, *Geocarto International*, 1-19.
- Yang, P., Y. Shang, Y. Li, H. Wang, and K. Li, 2017. Analysis of Potential Rockfalls on a Highway at High Slopes in Cold-Arid Areas (Northwest Xinjiang, China), *Sustainability*, 9(3): 414.
- Youssef, A.M., H.R. Pourghasemi, Z.S. Pourtaghi, and M.M. Al-Katheeri, 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia, *Landslides*, 13(5): 839-856.